

FAUXDIO: An audio deepfake detector for law enforcement and forensics

*Anil Alexander¹, Linda Gerlach¹, Thomas Coy¹, Oscar Forth¹,
Liam Lonergan² and Finnian Kelly¹*

¹Oxford Wave Research, Oxford, UK

{anil|linda|tom.coy|oscar|finnian}@oxfordwaveresearch.com

²Phonetics and Speech Laboratory, Trinity College Dublin, Dublin, Ireland.

llonerga@tcd.ie

Synthetic speech generation has now reached a level of sophistication allowing natural and realistic speech, capable of easily fooling the human ear, to be generated with relative ease either through text-to-speech synthesis or voice conversion. Linguistic and phonetic cues that could previously be relied upon as reliable indicators of fake content can now be masked to a certain extent using voice conversion with a suitable donor voice. Synthetic speech targeting a specific individual (often referred to as audio deepfakes) can now play a part in forensic cases including extortion and blackmail, false implication of innocent individuals, fraud and political disinformation. Development of countermeasures and tools to detect and defeat deepfakes has become the focus of governments, commercial institutions, and academia around the world. As part of the 2024 UK Home Office Deepfake Detection Challenge (Shanks, 2024), we developed an audio deepfake detection solution called FAUXDIO, which is capable of ingesting an audio or video file and automatically outputting an indication of whether the speech contained within the file is potentially real or fake. FAUXDIO relies on a DNN (deep neural network) detection model trained with many examples of real and fake speech and has a configurable decision threshold. The detection model can be run fully offline (e.g., on-premise) within a Windows desktop application for audio extraction (from video), conversion, and playback. FAUXDIO Desktop version, aimed at forensic users, enables user control via selectable detection models and configurable RAG (Red Amber Green) decision thresholds. The same detection models can also be run within FAUXDIO Web, which is a collaborative tool for multiple analysts to leverage insights obtained from samples provided by many users.

We calibrated the system using publicly available deepfake data including real deepfakes ‘in the wild’, as well as our internal test sets to provide meaningful RAG ratings. The tool further allows fine-tuning of decision thresholds and selection of detection models for specific use cases. We also developed a partial fake detection functionality which would highlight to the user potential partial fake regions in an otherwise largely real file. FAUXDIO Web provides an overall fake rating and score for an input file, along with a transcription of the speech, which is colour-coded to indicate which (if any) regions may be fake. Furthermore, the deepfake detector was connected with our MADCAT audio fingerprinting tool (Alexander et al. 2015) to recognise previously-seen fakes, thereby leveraging the ‘wisdom of the crowd’.

To demonstrate the detection performance on controlled data, we tested a subset of 1000 samples from the ASVspoof5 dataset (Wang et al. 2025) (780 fake samples, involving 9 different fake speech generation algorithms, and 220 real samples), resulting in an EER of 0.12%. Figure 1 shows a scatter plot of the data with fakes as blue triangles and reals as red circles. The FAUXDIO audio deepfake detection tool (see Figure 2) aced the 2024 Deepfake Challenge and has demonstrated similar strong performance in subsequent 2025 Home Office office user-trials and benchmarking.

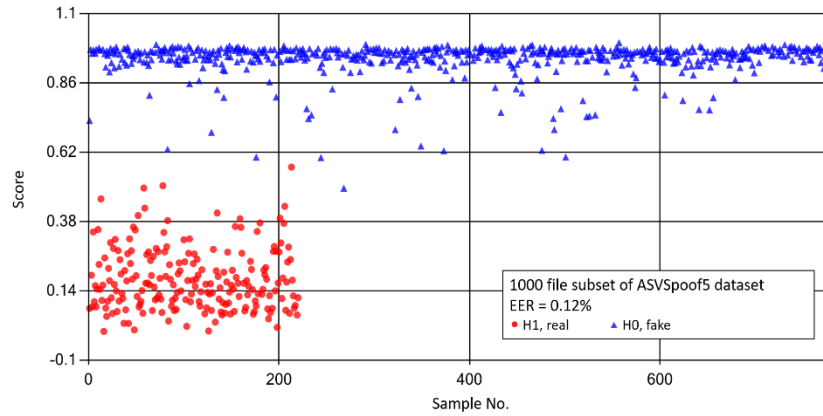


Figure 1. Scatter plot of FAUXDIO results based on ASVspoof5 data.

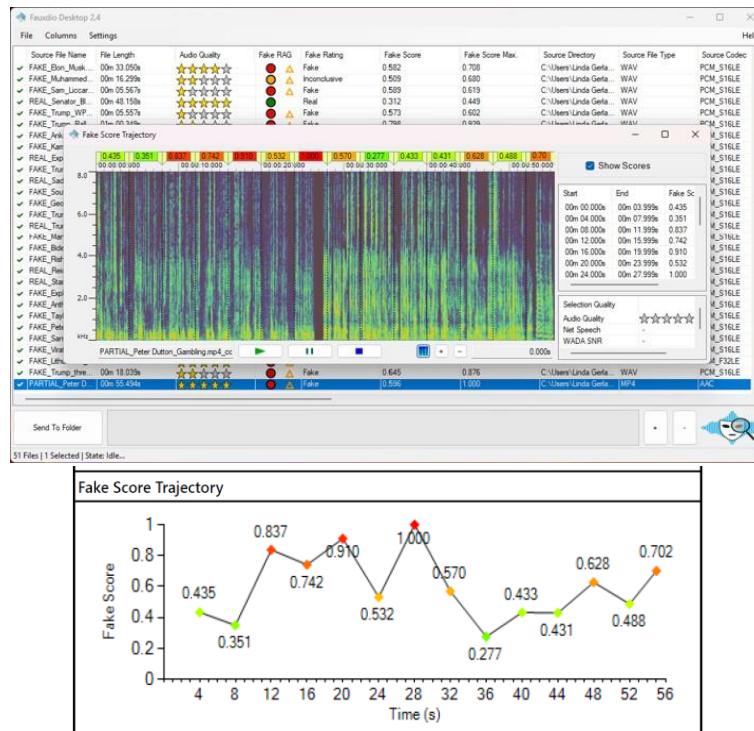


Figure 2. FAUXDIO Desktop interface with fake score segments highlighted in red for a partially fake audio file (above) and its fake score trajectory from the generated report (below).

References

- Accelerated Capability Environment (ACE; 2025, February 05). Innovating to detect deepfakes and protect the public. *UK government case study*. <https://www.gov.uk/government/case-studies/innovating-to-detect-deepfakes-and-protect-the-public>
- Alexander, A., Forth, O., Atreya, A. (2015). Audio fingerprinting to detect illegal content in digital media files. In *Proc. International Association for Forensic Phonetics and Acoustics (IAFPA) conference 2015*. Leiden, The Netherlands.
- Shanks, K. (2024, July 30). Innovative solutions unveiled at the Deepfake Detection Challenge Showcase. *Accelerated Capability Environment (ACE) blog*. <https://ace.blog.gov.uk/2024/07/30/innovative-solutions-unveiled-at-the-deepfake-detection-challenge-showcase/>, retrieved on 28.03.2025.
- Wang, X. et al. (2025), ASVspoof 5: Design, Collection and Validation of Resources for Spoofing, Deepfake, and Adversarial Attack Detection Using Crowdsourced Speech, *arXiv preprint*. <https://arxiv.org/abs/2502.08857>